SOME ECONOMICS OF AIR POLLUTION-INDUCED CHRONIC ILLNESS*

by

Shaul Ben-David
University of New Mexico

and

Thomas D. Crocker and Shelby D. Gerking
University of Wyoming

* This research was supported by the United States Environmental Protection Agency.

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

I. INTRODUCTION

At the time of the national awakening about environmental issues that occurred in the late 1960's, a great deal of public and scientific attention was focused on statistical relationships between air pollution and human health. While this research was undertaken with a large measure of academic curiosity, a major impetus was provided by Federal government agencies, such as the United States Environmental Protection Agency and its predecessors. The motivating factor for this agency encouragement was a laudable desire to establish scientific evidence for regulations designed to mitigate any detrimental health consequences of air pollution. For a time in the mid-1970's, the subject, though continuing to be discussed in scientific councils, did not capture much public attention, perhaps because of substantial reductions in the ambient concentrations of several common air pollutants. However, with the immediate threat that switching from oil and natural gas to coal fuels poses to the progress of a decade in controlling air pollution, the aforementioned statistical relationships are again a subject of public as well as scientific scrutiny.

In this paper, we assess the extent to which exiting epidemiological research can be interpreted as statistically demonstrating a relationship between air pollution and human health status. We also present some additional statistical research of our own. The next section is a critical review of the methodological underpinnings of existing research in air pollution epidemiology. So as not to exempt our previous work from this critical

review, we devote a third section to self-appraisal. A fourth section presents some new empirical results meant to respond to several of the faults we confessed in the third section. The two concluding sections summarize what we think we have thus far learned and make some suggestions for future research.

II. A CRITICAL REVIEW OF OTHERS' WORK

Much of the recent work in air pollution epidemiology has focused upon estimation of some version of the following expression:

$$H_i = a + bP_i + cX_i + u_i, \tag{1}$$

where H is a measure of morbidity or mortality, P is a measure of pollution, X is a set of other variables thought to influence health status, u is an error term that captures the effects of unmeasured influences upon health status, i indexes the individuals or groups of individuals in a sample, and a, b, and c are parameters to be estimated. Epidemiological work of this sort, a large part of which has been done be economists, presumes that there exists a distribution across individuals of tolerances to air pollutants and that there exist some individuals for whom any air pollution exposures whatsoever will trigger a decline in health status. This prespective may be contrasted with another, common to many epidemiological studies originating in the biomedical disciplines and sanctified in existing Federal clean air legislation, which posits a positive level of air pollution below which no individual will suffer a decline in health status[1].

Two recent empirical applications of the latter perspective are Morris, et al. (1976) and Bauhuys, et al. (1978). inspired by the principles of experimental design, the researchers in each of these studies selected two communities similar in most respects other than air pollution. Using analy-

sis of variance techniques, statistically significant differences in health status between the populations of the communities were then sought. Whether or not these differences were found, toxicological evidence from laboratory studies was then cited to provide a basis for rejecting or failing to reject air pollution as a cause of the difference. Many of the cited laboratory studies are, in principal, structured in the same fashion as the epidemiological studies; that is, the experimenter takes a treatment group and a control group of similar individual organisms and increases the pollution exposures of the treatment group until a decline in health status is observed. The pollution level at which this decline is first observed is then said to be the threshold at which pollution is universally unhealthy. Practitioners of this perspective generally agree that most substances commonly termed air pollutants can have deleterious human health effects. The controversies among them erupt over the threshold pollution levels at which these effects emerge and whether these threshold levels are found in everyday human environments. Because the methods provide no information on the magnitudes of any effects that do exist, the controversies are limited to questions on the statistical determination of the existance of an effect.

Unless all factors that contribute to differences in health status across individuals and locations can be controlled, the weaknesses inherent in empirical applications of the above perspective are apparent. In particular, statistically significant differences between the health states of two groups of individuals may not be observable because the contributions of air pollution to the true differences are overwhelmed by uncontrolled factors. Any perceived threshold is then more a matter of experimental design rather than of effect: perception of where the threshold lies will differ with the

extent to which the investigator is initially able to make his samples identical in all but their air pollution exposures. Moreover, even if the samples are identical, the outside observer gets the strong impression that there exists great confusion about the criteria for experimental design, the physiological and metabolic responses that consitute excess health impacts, the validity of extrapolating from animals to humans, and the processes that generate any defined health impact[3].

As is well known, the multivariate regression procedures usually used by economists investigating the health effects of air pollution allow explicit discrimination between the effects of air pollution, the effects of other observed control factors, and the effects of unobserved, presumably random factors. Although the estimated health effects of pollution will be biased if some of the assumed random factors vary systematically with pollution, the continuous covariation between health states and pollution that the procedures permit does not force one to adopt the ambiguous notion of a human health effects threshold before research is even initiated[4]. Neither is the investigator put in the uncomfortable position of having to assign the residual ("excess" deaths or illnesses) to something particular such as air pollution.

The first attempt to investigate the health effects of air pollution at a national level without the resumption of a threshold was the pathbreaking effort of Lave and Seskin (1970). Using 114 U.S. metropolitan areas as units of analysis, they employed single equation, ordinary-least-squares methods to regress 1960 mortality rates linearly upon ambient concentrations of sulfates and particulates, and other plausible influences upon mortality. They tentatively concluded that statistically significant health effects of air pollution existed. This original study has inspired a substantial number

of similar subsequent studies, including the culminating effort of Lave and Seskin (1977)[5]. Without exception, all have discerned a close and substantial inverse association between mortality rates and one or more air pollutants. Recently however, two studies have become available that should give considerable pause to those wishing to accept the Lave-Seskin, et al. findings.

Smith (1977), using data for 50 U.S. metropolitan areas in 1968-1969, applied versions of the Ramsey (1969) tests for specification error in the general linear model to 36 different single equation specifications. These specifications were similar, and often identical, to those greeted with the most approval by the authors of the Lave-Seskin, et al. literature. None of the specifications could pass all of the Ramsey (1969) tests at the 10 percent level, although four passed all tests except that for non-normal errors.

The Ramsey (1969) tests are meant to be used to assess conformity with the basic assumptions for error structure of the classical linear model. They give no hint about events when attempts are made to correct for one or more of the specifications errors. In a recent paper, Cracker-Schulze, et al. (1979, pp. 24-71) use 1970 mortality data from 60 cities while trying to correct for potential omitted independent variable and simultaneous equation problems. Upon adding measures of medical care, cigarette consumption, and diet to the single equation Lave-Seskin, et al. specifications, they found no statistically significant effect of nitrogen dioxide, total suspended particulates, and sulfur dioxide upon the rate of total mortality[6]. Retaining the former variables, and accounting for the plausible simultaneity between health status and medical care, did nothing to improve the statistical sign-.

ificance of the three air pollution variables. On the presumption that these findings were sufficient to demonstrate the weakness of the Lave-Seskin type results, the authors did not go on to account for the obvious simultaneity between median age (or percentage over 65 years) and mortality incidence, income and mortality incidence, and several other plausible sources of simultaneity.

The results obtained by Smith (1977) and Cracker-Schulze, et al. (1979) cast doubt upon the robustness of the Lave-Seskin, et al. estimates, in spite of the no-threshold perspective embodied in these estimates. Nevertheless, before dismissing the hypothesis of an inverse relation between everyday air pollution levels and health states, it must be recognized that Lave-Seskin, et al, may have been asking more of their data than it was capable of giving[7]. Less than one in every 100 people dies in the U.S. each year. No biomedical authority asserts that air pollution is the dominant cause of the deaths that do occur. Many take the view that it is the direct cause of no more than a small fraction of these deaths, although they would agree that it may be quite important in intensifying predispositions toward mortality. However, the general properties of the underlying processes that encourage this predisposition are ill-understood. Thus, even with quite large samples, available estimation techniques and a priori knowledge may be inadequate for distinguishing the mortality effects of air pollution in a human population sample from a host of similar and plasible minor contributing factors.

The possible inadequacy of many available techniques for estimating the existence and/or magnitude of air pollutant-induced mortality applies with special force, given the data Lave-Seskin and their successors had to employ. Their work can be interpreted as an attempt at establishing the probability

of a representative individual currently residing in a representative region dying in a given year from a geographically representative level of air pollution occuring in a representative year. Since they had no information about the distribution of influential health factors, including air pollution, across the urban areas constituting their units of analysis, the identifying variabilities of their samples were perhaps drastically reduced. [8] When this relatively low variability of the samples is coupled with what are probably substantial measurement errors in the air pollution variables, the baggage of additional explanatory variables and more sophisticated estimation techniques to correct for specification error that the data are able to carry must be rather light. The attempted corrections may serve only to misinform. Furthermore, that which is being corrected may be only an apparition since, as Crocker (1975, pp. 350-351) demonstrates, the measure of (the probability of) death, employing some, group of individuals as the fundamental unit of observation, can differ from one group to another; there could be as many unique measures employed as there are groups.

The preceding remarks lead us to three conclusions. First, given the biomedical and economic subleties inherent in comprehending the etiologies of air pollution-induced mortality and morbidity, the estimates obtained from aggregated data used in the great bulk of extant studies are unlikely ever to be sufficiently compelling to establish a consensus. Only the use of actual individuals as fundamental units of observation is likely to provide enough strangth in the data base to.carry the requisite statistical burdens. Second, the statistical burdens that have to be carried might be considerably lightened if research concentrates on borbidity rather than mortality. The

frequency, and most likely the identifying variability, of the former is greater by a factor of fifteen or twenty. Finally, because one's health status is influenced by the choices one makes about lifestyles, environmental and occupational exposures to possible toxics, and other health-influencing factors, economics can provide a priori hypotheses and an analytical framework to lend additional structure to epidemiological investigations. The relationships with which observed real world outcomes are consistent can, therefore, be further narrowed.

III. A CRITICAL REVIEW OF OUR WORK

Crocker-Schulze, et al., (1979) embodies both mortal ity and morbidity studies. The mortality study had the essentially negative purpose of empirically demonstrating that the estimates derived in Lave-Seskin type studies are not at all robust. The morbidity study had the more positive purpose of investigating air pollution and human health stutus with a data set better able to bear added statistical burdens and to accept hypothesis testing about the impact of man's free will upon health status. In this section, we briefly discuss several entirely correct ways in which the morbidity study is susceptible to injury. Strangely, although the study has been carefully persued by many interested parties, few have hit it where their thrusts could not even begin to be countered without additional work on our part. Here, we present some of those thrusts.

Depending almost entirely upon ordinary-least-squares (OLS), the morbidity study estimated the effect of air pollution upon self-reported health status measured as length of time chronically ill and annual frequency of acute illnesses. Expressions linear in the original variables were estimated for

several 400 person samples independently drawn from all household heads in the Panel Survey of Income Dynamics (PSID) [Survey Research Center (1972)] who had always lived in one state. Although some attention was devoted to NO_2, air pollution was generally measured as the annual 24-hour geometric mean of SO_2 and/or TSP in the head's county of residence for the year (1967-75) from which the sample was drawn. In addition to air pollution, measures of the intensity of the head's illness, his biological and social endowments, life-style, and work, home, and outdoor environments were, when available, included as explanatory variables. Air pollution contributed positively and significantly to both chronic and acute illnesses in the majority of the unpartioned samples. Upon combining these dose-response estimates with a simple recursive labor supply formulation, the economic impact of air-pollution-induced chronic illness upon labor productivity was estimated to exceed that of air pollution-induced acute illness by nearly a factor of 20.

These results encouraged us to proceed further, particularly with respect to investigating air pollution-induced chronic illness. The obvious initial further step was to correct some of the outstanding technical problems in our treatment of the dose-response functions estimated from the PSID data. 9/ These problems fall into three general categories: (1) the definition of self-reported health status; (2) the factors used to explain self-reported health status; and (3) the algorithm used to estimate self-reported health status.

The PSID data on the chronic illness health status of household heads consists only of responses to four questions stated in the following order:

1. Do you have a physical or nervous condition that limits the type of work you can do or the amount of work that you can do?

2. How much does it limit your work?

3. How long have you been limited in this way by your health?

4. Is it getting better, worse, or staying about the same?

In the case of the first question, persons were asked for a yes or no answer, while for the remaining three questions the response called for was categorical. The response to question #3 was used as the dependent variable in our earlier analysis. However, the responses to this question were recorded categorically with the uppermost category being bounded only by age. Moreover, this response was conditional upon the response to question #1 and possibly question #2. For these reasons, interpretation of the earl ier chronic illness dose-response estimates required a string of assumptions that may or may not have been important to stated results. In any case, in order to assess the validity of the earlier results, it is prefereble to remove any clouding that the assumptions may have introduced. The response to question #1 is unambiguous.

Even though the response to question #1 is unambiguous in terms of self-reported health status, it need not represent the respondent's clinical health status. More specifically, individuals may not be alike in the way they determine whether or not they are chronically ill. Economic factors including type of job, access to disability benefits, and other measures of the opportunity costs of not working may be important to this determination. For example, consider two persons who are alike in every respect other than their hourly wage. The person with the lower of the two wage rates will have a lower opportunity cost of not working. He may be perfectly healthy but desire to work fewer hours and use illness as an excuse, or he may actually be sick more often than his higher income counterpart because he does not find it economically advantageous to be as healthy.

The preceding suggests that our earlier estimated chronic Illness dose-response expressions might be biased because economic determinants of self-reported health status were omitted. In addition to these economic determinants, other, more traditional life-style, biological endowment, medical care, and environmental determinants were omitted or imperfectly measured. For example, the earlier estimates included no information on job accident rates, and used cigarette expenditures as an index of cigarette consumption. These variable exclusions and imperfectly measured explanatory variables can bias the estimated contribution of air pollution to self-reported health status.

Finally, given the chronic illness health status variable employed in our earlier work, the use of an OLS estimation procudure could have been inappropriate for two reasons. First, self-reported health status might have been determined jointly with some explanatory variables (e.g., liesure exercise, cigarette smoking, and medical care) that were also choice variables. OLS estimates of the chronic illness dose-response expression would then be biased and inconsistent. Second, the health status variable was recorded in a categorical rather than in a continuous fashion. This means that heteroskedasticity could be present in the OLS-estimated chronic illness dose-response expressions with a consequent introduction of biases in the standard errors of the air pollution coefficients. As McKelvey and Zavoina (1975) show, the use of OLS procedures with categorical dependent variables can cause the relative impacts of certain variables to be severly underestimated.

IV. SOME NEW, BUT STILL LIMITED RESULTS

In this section, we present some new results which, insofar as available data allow, correct partially or wholly for the technical problems raised in the previous section. The outstanding failing of these new results is that we do not construct an explicit analytical model to account for the economic determinants of self-reported health status. Instead, we do no more than introduce explanatory variables such as family assets and union membership that would plausibly have a role to play in expressions derived from any analytical model dealing with the effect of the opportunity costs of not working upon perceived own health status.

Table 1 lists the variables we employ. Alcohol expenditures, numbers of daily cigarettes smoked, free access to medical care, physician population, carcinogenic potential in the workplace, precipitation, workplace job accident rate, current transfer income, and union membership all represent variables that did not appear in our previous chronic illness dose-response expressions. Separate structural expressions are estimated for numbers of daily cigarettes smoked, whether or not the individual has medical insurance, and whether or not he participates in strenuous liesure exercise on the presumption that they are jointly determined with health status. To account for plausible nonlinearities with respect to the impact of age and food expenditures on health status, squared, as well as original, values are entered for these variables.

In view of the categorical nature and the simultaneity of the dependent variable, the estimation technique selected was the two-stage limited depen-dent variables (2SLDV) approach suggested by Nelson and Olson (1978). More specifically, the estimation procudure these authors propose is to:

TABLE 1

COMPLETE VARIABLE DEFINITIONS

Self-Reported Health Status Variables

DSAB - Limitation on work = 1; otherwise = 0

LDSA - Disabled for $\leq$ 2 years = 1; 2-4 years = 2; 5-7 years = 3; $\geq$ 8 years = 4; otherwise = 0.

Biological and Social Endowment Variables

AGE - Age in years.

EDUC - Completed. 6-8 grades = 2; 9-11 = 3; 12 grades = 4; 12 grades plus non-acedemic training = 5; college, no degree = 6; college degree = 7; advanced or professional 'degree = 8; otherwise = 1.

FMSZ - Family size in number of persons in housing unit.

POOR - Stated that parents were poor "...when you were growing up..." = 1; otherwise = 0.

SEX - Male = 1; Female = 0.

Lifestyle Variables

ALKY - Annual alcohol expenditures X 10^{2} per adult family member.

CIGN - Number of daily cigarette packs smoked per adult family member. This variable was calculated by dividing the PSID data on 1970 cigarette expenditures by the 1970 retail price of a pack of cigarettes in the 1970 state of residence. Retail price data was taken from Tobacco Tax Council, Inc. (1978, pp. 67-69).

FOOD - Family food consumption relative to food needs standard in percent. Consumption refers to food expenditures in dollars and includes amounts spent in the home, school, work, and restaurants, as well as the amount saved in dollars by eating at work or school, raising, canning, or freezing food, using food stamps, and receiving free food. The food needs standard is in dollars and is based on USDA Low Cost Plan estimates of weekly food costs as published in the March 1967 issue of the Family Economics Review. The standard itself is calculated by multiplying the aforementioned weekly food needs by 52 and making a series of adjustments according to family size.

LEXR - Indication that dominant leisure-time activities involves strenuous exercise = 1; otherwise = 0. Strenuous activities were said to include fishing, bowling, tennis, camping, travel, hunting, dancing, motorcycling, etc.

Health Care Variables

HVET - Free access to medical care as a veteran or through medicaid = 1; otherwise = 0.

INSR - Has hospital or medical insurance = 1; otherwise = 0.

PHYS - Physicians per 10,000 population in county of residence on July 1, 1975. This data was obtained from U.S. Bureau of the Census (1978, Table 2).

Environmental Variables

CANX - An index of workplace "carcinogenic potential" by two-digit SIC code as presented in Hickey and Kearney (1977) and determined by dividing their Table 8 by their Table 7. We are aware that these authors insist that "...the magnitude of the derived carcinogenic potential is not suitable for any health hazard inference" (p. iii).

COLD - Mean annual January temperature in the 1970 county of residence in F° X 10. This data is from U.S. Bureau of the Census (1978, Table 4).

PRCP - Mean annual precipitation in inches X 10^2 in the 1970 county of residence. This data is from U.S. Bureau of the Census (1978, Table 4).

JACCR - Number of disabling work injuries in 1970 by 2 and 3-digit SIC code for each million employee hours worked. The data is from Table 163 of Bureau of Labor Statistics (1972).

SULM - Annual 24-hour, geometric mean sulfur dioxide micrograms per cubic meter as measured by the Gas Bubbler Pararosaniline-Sulfuric Acid Method. The data were obtained from the annual USEPA publication, Air Quality Data - Annual Statistics,and refer to a monitoring station in the 1970 county of residence.

TSPM - Annual 24-hour geometric mean total suspended particulates in micrograms per cubic meter as measured by the Hi-Vol Gravimetric Method. The data were obtained from the annual USEPA publication, Air Quality Data - Annual Statistics, and refer to a monitoring station in the 1970 county of residence.

Pecuniary Variables

ASSETS - Sum of 1970 income in dollars X 10^2 from social security, retirement 'pay, pensions, annuities, dividends, interest, and rent.

UNION - Member of a labor union = 1; otherwise = 0.

(i) Estimate the reduced form of the structural system by applying an appropriate maximum likelihood technique to each.

(ii) Form instruments from the "predicted" values of the dependent variables using the observations from the sample on the exogeneous variables together with the estimated reduced from coefficients obtained in the first step.

(iii) Replace the jointly dependent variables on the righthand side of the equations in the structural system with their -instruments constructed in the second step.

(iv) Estimate the resulting relations by an appropriate maximum likelihood method.

As can be easily seen, this estimation procedure applied to a system of simultaneous equations is just two-stage least squares in the case where all jointly dependent variables are continuous over the entire real line. However, the approach of Nelson and Olson (1978) takes account of the fact that some dependent variables, particularly the DSAB variable of interest here, do not exhibit this type of behavior. They therefore suggest that an appropriate limited dependent variable technique be used in the estimation of both the reduced form and the structural form of the model. In this case, since DSAB is defined to take on only the values of zero or one, the probit model would appear to be the most appropriate of the alternative limited dependent variable methods.

The procedures outlined above were applied to a sample of 309 individual household heads drawn from the 1970 calendar year of the PSID sample. All individuals had always resided in the 1970 state of residence. We are, thus, able to control partially for the air pollution exposure history of the individual, given that relative 1970 pollution concentrations across residential locations are similar to the history of relative concentrations. The

year 1970 was selected for detailed empirical analysis because the chronic illness dose-response expressions estimated for this year in Crocker-Schulze, et al. (pp. 105-109) were considered to be the best representatives of all the expressions for assorted years estimated by ordinary-least-squares from the PSID data.

The 309 individuals of the sample represent all individuals in the 1970 PSID calendar year data for whom we were able to obtain observations on each explanatory variable, including total suspended particulates and sulfur dioxide. It should be noted that this sample is unlikely to correspond to a random sample of the U.S. population. If anything, as a glance at the arithmetic mean values of the explanatory variables presented in Table 2 shows, the sample appears to include a somewhat disproportionately high number of female household heads, "poor" childhood backgrounds, and relatively low pecuniary values of family assets. For our present purposes, of course, a random sample is unnecessary, given that the sample was not selected on the basis of whether or not the individual reported he suffered from a chronic illness.

The results of estimating the augmented (relative to our previous work) chronic illness dose-response expression by the multivariate Probit estimator are reported in the last two columns of Table 2. As Poirier and Melino (1978) have shown, the coefficients are proportional to the change in the probability that an individual will report being chronically ill for a one unit change in the explanatory variable. Thus, for example, a male, is nearly twice as likely to report being chronically ill as is a female. Our use of the Probit esitmator presumes that each individual has a threshold level of the explanatory variable below which he will not view himself as being made chronically

ill. However, the estimator also presumes that there exists a transformation causing these threshold values to be normally distributed over our sample and, therefore, that there exist some individuals for whom even minor levels of air pollution will cause them to report being chronically ill. The constant term is simply a shifter.

With the exceptions of CIGN, LEXR, and POOR, the signs of all coefficients coincide with a priori expectations. The combinations of signs for the AGE variables and the FOOD variables are consistent with increased likelihoods of reporting chronic illness at the extremes of age and diet adequacy with a reduced likelihood in the middle ranges. increases in alcohol consumption, exposures to carcinogenic substances, accident risks in the workplace, physicians to originate or confirm the individual's self-diagnosis, and air pollution in the form of sulfur dioxide all serve to increase the chances of self-reported chronic illness. The coefficients of CANX and JACCR are probably biased downward, since they refer only to the current workplace, rather than to the individual's workplace history. On the other hand, consistent with the work of Tromp (1962) and others, high precipitation and low midwinter temperatures are less likely to make the individual feel chronically ill. Those variables such as ASSETS and UNION, represnting factors thought to reduce the opportunity costs of feeling chronically ill, all contribute positively to the probability of reporting chronic illness. Similarly, more education and larger family size, variables which capture factors tending to increase the opportunity costs of feeling chronically ill, each have negative signs attached. Since people who are vetrans and have medical insurance face lower marginal prices for medical care, they can be expected to consume more medical care thereby reduce the frequency of their

chronic illnesses. The negative signs attached to HVET and INSR are consistent with this interpretation. Note that the coefficient attached to the latter variable is estimated from a system that accounts for the simultaneity between the likelihood of possessing medical insurance and the presence of chronic illness. Note also, however, that the results for these variables explaining the "demand" for chronic illness have not been derived from an explicit analytical model. The above interpretation may therefore be unwarranted.

Interpretations for. the signs of CIGN, LEXR, and POOR are less readily provided. It is possible that no one of these variables is a reasonable measure of the effect we were trying to capture. For example, CIGN represents the estimated number of current cigarettes smoked per adult family member. There is no obvious connection between this measure and the smoking history of the individual whose health status is being inspected. It is, of course, possible that those who are already chronically ill increase their smoking because of the, greater utility it might then afford. As for LEXR, it appears from its estimated mean value that the expression used to calculate it did not perform very well. In addition, the preception of what constitutes strenuous exercise can differ across individuals Again, strenuous exercise might yield greater utility for those who are already chronically ill, so that they are more likely to participate in it than are healthy individuals. Similarly, the current perception of whether one's parents were poor may be more a measure of one's current real income status relative to the former status of one's parents rather than an absolute measure of the latter's former status. Thus, extending the Dusenberry (1949) hypothesis to an intergenerational context, it might be that greater relative

TABLE 2

MAXIMUM LIKELIHOOD ESTIMATES OF SELF-REPORTED CHRONIC ILLNESS (DSAB)

Variable		Mean	Coefficient	Standard Error
AGE		39.36	0.084	0.054
$(AGE)^2 \times 10^{-1}$		177.00	-0.776	0.582
ALDY		1.11	0.169	0.100
ASSETS		2.68	0.001	0.001
CANX		18.77	0.006	0.021
$\widehat{CIGN}$	$CIGN =$	1.73		
	$\widehat{CIGN} =$	0.64	-0.527	0.190
COLD		37.86	-0.025	0.015
EDUC		3.76	-0.087	0.162
FMSZ		3.22	-0.005	0.056
FOOD		1.80	-0.499	0.470
$(FOOD)^2$		3.90	0.089	0.095
HVET		0.19	-0.472	0.400
$\widehat{INSR}$	$INSR =$	0.72		
	$\widehat{INSR} =$	0.80	-1.223	0.430
JACCR		33.17	0.003	0.005
$\widehat{LEXR}$	$LEXR =$	0.18		
	$\widehat{LEXR} =$	-1.13	0.115	0.454
PHYS		24.08	0.007	0.010
POOR		0.52	-0.503	0.290
PRCP		39.77	-0.043	0.017
SEX		0.57	0.927	0.556
SULM		18.37	0.011	0.010
UNION		0.19	0.422	0.398
Constant			1.090	1.807
(-2.0) times log of likelihood ratio			85.609; statistically significant at the one percent for the χ^2 distribution with 21 degrees of freedom.	
Observations at Unity			77	
Observations at Zero			232	

NOTE: No levels of significance are indicated because the asymptotic properties of the standard errors for this sample are not known. A simulation experiment with the simultaneous probit estimator suggested to Nelson and Olson (1978, p. 702) that its standard errors could be biased upward by as much as a factor of 1.6.

current real income may engender a sense of security reducing the opportunity costs of being chronically ill. Alternatively, the explanation for the unexpected negative sign might simply be that a selection process operated in the past to eliminate those who were less well genetically endowed and who also had poor childhoods.

A rank-ordering of the explanatory variables from the most to the least statistically significant results in the following: CIGN, INSR, PRCP, POOR, ALKY, SEX, COLD, AGE, $(AGE)^2$, HVET, FOOD, UNION, SULM, ASSETS, $(FOOD)^2$, PHYS, JACCR, EDUC, CANX, LEXR, AND FMSZ. Thus, at least for the sample represneted in Table 2, air pollution, as measured by annual 24-hour geometric mean sulfur dioxide, is less robust statistically than the climate variables but more robust than the measures of occupational hazards. However, as indicated in the table, SULM would appear to be statistically insignificant at conventional levels. This general conclusion holds when another air pollution variable, annual 24-hour geometric mean suspended particulates, replaces the measure of sulfur dioxide used in Table 2. Upon doing this, a coefficient of 0.006 with a standard error of 0.007 is obtained. Given that the standard errors of the simultaneous probit estimator are thought to be biased upward (perhaps by as much as 1.6 according to Nelson and Olson (1978, p. 702), the actual effect of air pollution on self-reported health status may be more significant than our results indicate. Nevertheless, even if the standard error on the air pollution coefficients are in fact biased upward by a factor of 1.6, the statistical significance of these coefficients remains questionable.

In order to provide another basis for comparison with Crocker-Schulze, et al. (1979), we substituted the measure used for the length of chronic

illness (LDSA) in our earlier work for the dependent variable in Table 2. The system was estimated by the two-limit simultaneous probit technique employed in Nelson and Olson (1978). Again, the results obtained were not inconsistent with our previous OLS estimates. In fact, the magnitudes of the air pollution coefficients were almost twice those obtained in the OLS results. However, as Poirier and Melino (1978) demonstrate, the coefficients of an explanatory variable in a truncated regression procedure such as probit is proportional to, but not equal to, the partial derivative of the conditional mean of the dependent variable with respect to a one unit change in an explanatory variable. This factor of proportionality, which is identical for each coefficient in a regression, can be determined when the variance of the untruncated variable is known. For the PSID data set, this variance is unknown.

V. WHITHER FROM HERE

The motivation for this paper, as well as our previous work in the area, originated in our convictions that economic analysis and its empirical techniques could contribute to the resolution of certain recurring puzzles in studies of the incidence and severity of diseases in human populations, particularly the epidemiology of air pollution. We have viewed human health status as a decision variable and have therefore been able to employ economic theory as a means of providing more a priori structure for the analysis of epidemiological data. Considering only the empirical results reported in the previous section, it seems we have not yet provided enough information on structure for resolution. We have by no means, however, exploited all the conceivable economic-behavioral structural relations from which restrictions might be obtained.

One might introduce more statistical information by quasi-replication of the structures already estimated; that is, we could pull additional samples from the PSID data set and estimate for each of those samples the same two structures already discussed. This strategy has been used [Crocker-Schulze, et al. (1979)] in an earlier substantially less rigorous treatment of the same data.

Alternatively while retaining the structure that economic analysis and epidemiology provide, we can draw upon knowledge in biophysics, biochemistry, and bioenergetics to a much greater degree than previous studies in air pollution epidemiology appear to have done. In a manner consistent with human capital theory, as some existing work has in fact already done [e.g., Cropper (1977) and Crocker-Schulze, et al. (1979)]. The individual might be construed as having an initial health endowment that, due to natural aging, depreciates exogenously over time. However, by his decisions about life-style and his occupational and environmental exposures, he can either slow or accelerate this natural depreciation. An integral part of these human capital treatments has been the representation of a production function in implicit form where some crude measure of health status is determined by rather arbitrary assortments of the aforementioned collection of life-style, occupational, and environmental variables. We suggest, at least insofar as empirical treatments are concerned, that one can specify this production function in much more detail while retaining the human capital framework for the individual's decision problem.

As an alternative to traditional toxicological research emphases upon metabolities and metabolic pathways, the Second Task Force for Research Planning in Environmental Health Science (1977, Chapter 14) recommends that more

effort be devoted to building upon existing knowledge of the structure and function of particular organ systems such as the respiratory and cardiovascular systems. Contrary to most of the arcane (to an economist) basic research on the fundamental chemical processes at work in various metabolic pathways, much of the work on the determinants of the individual's research of organ function appears to be readily translatable into mere displays of the fact that within limits the same quality of some simple measure of the health status of the organ system, such as the ventilation capacity of thelung, can be obtained from various combinations of inputs[10]. In many cases, the responses of the health indicator of the organ system to various stresses follow well-known physical laws having specific functional forms and even particular values attached to coefficients.[11].

When writing down the individual's decision problem with respect to health status, we may be able to structure the problem more tightly by building the aforementioned information on organ system responses directly into the constraint set. Rather than having an implicit production function in which the value of a self-reported, highly aggregated measure of health status (e.g., whether or not the individual is chronically ill) is explained by a collection of intuitively reasonable variables, one can employ a description that precisely maps a limited and well-defined set of major influential factors into a continuous scaler measure of the health of an organ system.

V. SUMMARY AND CONCLUSIONS

The preceding pages are not without technical sin. In particular, with out rigorously explaining from whence they come, we have introduced variables that are supposed to represent the opportunity costs of reporting or failing to report ones self chronically ill. Otherwise, however, by employing a

more robust estimation procedure, by redefining the chronic illness variable, and by introducing better measures of cigarette smoking, hazards and toxic exposures in the workplace, medical care, and climate, we have responded to several well-founded criticisms of the morbidity results in Crocker-Schulze, et al. (1979). On the basis of those new tests, we see no reason to alter our previous interpretation of the effect of air pollution upon self-reported chronic illness.

FOOTNOTES

[1] In accordance with the eloquent argument of Calabresi and Bobbit (1978), one might attribute the dominance of this perspective in pubic policy settings to the fictions erected by societies to segment markets that would otherwise require explicit judgments about the relative worths of individuals' lives. Calabresi and Bobbit (1978) argue that these fictions serve to soften intolerable societal stresses. The purpose they serve in a scientific setting is not obvious.

[2] Alternatively, the laboratory studies try to specify the intervening processes causing an observed health effect.

[3] Apart from these issues, the practice of applying laboratory results to everyday human environments is questionable. As Anderson and Crocker (1971, p. 146) note, so as to remove all sources of stress other than air pollution, all other factors influencing health in the laboratory tend to be set at biologically optimal levels. Given that these biologically optimal levels exceed those found in everyday environments, it follows from the law of variable proprotions that air pollution-induced health effects in the laboratory will exceed those found in everyday environments.

[4] It should be noted that many biomedical authorities strongly dispute the biological existence and the policy relevance of thresholds for most environmental contaminants. Authors such as Epstein (1974), Goldsmith and Friberg (1977) argue that any positive amount of pollution induces ill-health effects for some individuals and increases the probability of ill-health for everyone exposed.

[5] Among the more notable examples are: McDonald and Schwing (1973); Liu and Yu (1376); Mendelsohn and Orcutt (1979); Gregor (1977); and Koshal and Koshal (1973).

[6] However, particulates was statistically significant in an expression explaining pneumonia and influenza related deaths. Sulfur dioxide was statistically significant in an expression for deaths attributed to early infant diseases. Nitrogen dioxide would have been statistically significant in heart disease if a slightly less severe level of acceptance had been adopted.

[7] In order to get the data to "give" more, the authors of the Lave-Seskin type work have usually tested with the same data set several different functional forms and combinations of explanatory variables. The objective frequently seems to have been the maximization of certain summary statistics (e.g., the coefficient of determination) having no basis in any a priori hypothesis. We are unaware that the pretest or selection procedures surveyed Wallace (1977) and Judge, et al. (1980, Chap. II) have ever been employed during these manipulations. If these procedures are not employed, the properties of the classical least squares estimators these authors typically use can be substantially altered; that is, the customary interpretations cannot be attached to estimated coefficients and standard errors.

[8]Ambient pollution concentrations for a single year at single (usually downtown) sites served as proxies for the lifetime exposure histories of entire regional populations. For a succinct treatment of the trade-off between corrections for specification error and identifying Variability when measurement error is present in an independent variable of interest, see Griliches (1977, pp. 12-13). The addition of imperfectly measured explanatory variables to the expression being estimated will bias downward the coefficients of the air pollution variables.

[9]For now, we much prefer to leave accounting issues about what the estimate mean in terms of national economic impacts to more adventuresome types.

[10]See Kao (1972, Chap. III and IV) for readily understood treatments of the lung as a mechanical pump and as a gas exchanger.

[11]Many of these responses have been established in animal rather than human studies. The validity of extrapolating results from the former to the latter is a major source of controversy in biomedical studies of pollution effects upon organ systems.

REFERENCES

Anderson, R.J., Jr., and T.D. Crocker, "The Economics of Air Pollution: A Literature Assessment," in P.B. Downing, ed., Air Pollution and the Social Sciences, New York: Praeger Publishers (1971), 133-166.

Bouhuvs. A., G.J. Beck, and J.B. Schoenberg. "'Do Present Levels of Air Pollution Outdoors Affect Respiratory Health?" Nature 276 (Nov. 30, 1978), 466-471.

Bureau of Labor Statistics, Handbook of Labor Statistics, 1972, Bull. 1735, U.S. Department of Labor, Washington, D.C.: USGPO (1972),

Calabresi, G., and D. Bobbitt, Tragic Choices, New York: W.W. Norton (1978).

Crocker; T.D., "Cost Benefit Analysis of Cost-Benefit Analysis," in H.M. Pesken and E.P. Seskin, eds., Cost-Benefit Analysis and Water Pollution Policy, Washington, D.C.: The Urban Institute (1975), 341-360.

Crocker, T.D., W. Schulze, S. Ben-David, and A.V. Kneese, Experiments in Air Pollution Epidemiology, Washington, D.C.: USEPA Publication No 60015-79-001a (1979).

Cropper, M.L., "Health, Investment in Health, and Occupational Choice," Journal of Political Economy 85(Dec. 1977), 1273-1294.

Duesenberry, J.S., Income, Saving, and the Theory of Consumer Behavior, Cambridge, Mass. : Harvard University Press (1949).

Engel, G.L., "The Need for a New Medical Model: A Challenge for Biomedicine," Science 195 (Jan. 22, 1977), 129-136.

Epstein, S. "Environmental Determinants of Human Cancer, Cancer Research 34 (Oct. 1974), 2425-2435.

Goldsmith, J.R., and L.T. Fribert, "Effects of Air Pollution on Human Health," in A.C. Stern, ed., The Effects of Air Pollution, 3rd ed., New York: Academic Press (1977).

Gregor, J.J., Intra-Urban Mortality and Air Quality, Corvallis, Ore.: USEPA Publication No. 60015-77-009 (1977).

Griliches, A., "Estimating the Returns to Schooling: Some Econometric Problems," Econometrica, 45 (Jan. 1977), 1-21.

Hickey, J.L.S., and J.J. Kearney, Engineering Control Research and Development Plan for Carcinogenic Materials, Cincinnati, Ohio: U.S. Public Health Service under Contract No. 210-76-0147 (Sept. 1977).

Judge, G.G., W.E. Griffiths, R.C. Hill, and T. Lee, The Theory and Practice of Econometrics, New York: John Wiley and Sons (1980).

Kao, F.F., An Introduction to Respiratory Physiology, Amsterdam: Excerpta Medica (1972).

Koshal, R.K., and M. Koshal, "Environments and Urban Mortality; An Ecnonometric Approach," Environmental Pollution 4 (June 1973). , 247-259.

Lave, L.B., and E.P. Seskin, "Air Pollution and Human Health," Science 169 (Aug. 21, 1970), 723-733.

_____, Air Pollution and Human Health, Baltimore: Johns Hopkins University Press (1977).

Liu, B., and E. Yu, Physical and Economic Damage Functions for Air Pollutants by Receptor, Corvallis, Ore.: USEPA Publication No. 60015-76-011 (1976).

McDonald, G.C., and R.C. Schwing, "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics 15 (1973), 463-481.

McKelvey, R.D., and W. Zavoina, "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," Journal of Mathematical 'Sociology 4 (1975), 103-120.

Mendelsohn, P., and G. Orcutt, "An Empirical Analysis of Air Pollution Dose-Response Curves," Journal of Environmental Economics and Management 6 (1979), 85-106.

Morris, S.C., M.A. Shapiro, and J.H. Waller, "Adult Mortality in Two Communities with Widely Different Air Pollution Levels," Archives of Environmental Health 31 (1976)) 248-254.

Nelson, F., and L. Olson, "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables," International Economic Review 19 (Oct. 1978), 695-709.

Poirier, D.J., and A Melino, "A Note on the Interpretation of Regression Coefficients within a Class of Truncated Distributions," Econometrica, 46 (Sept. 1978)) 1207-1209.

Ramsey, J.B., "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," Journal of the Royal Statistical Society Series B 31 (1969), 350-371.

Second Task Force for Research Planning in Environmental Health Science, Human Health and the Environment: Some Research Needs, Washington, D.C. : USDHEW Publ. No. NIH 77-1277 (1977).

Smith, V.K., The Economic Consequences of Air Pollution, Cambridge, Mass.: Ballinger Publishing Co. (1977).

Survey Research Center, A Panel Study of Income Dynamics, Ann Arbor: Institute for Social Research, University of Michigan (1972).

Tobacco Tax Council, Inc., The Tax Burden on Tobacco, Washington, D.C.: Tobacco Tax Council, Inc. (1978).

Tromp, S.W., Medical Biometeorology, Amsterdam: Elsevier Publishing Co. (1962).

U.S. Bureau of the Census., County and City Data Rook, 1377, Washington, D.C.: U.S. Government Printing Office (1978).

Wallace, T.D., "Pretest Estimation in Regression; A Survey," American, Journal of Agricultural Economics. 50 (August 1977) 431-443.

THIS PAGE INTENTIONALLY LEFT BLANK